Double paths network with residual information distillation for improving lung CT image super resolution

Yihan Chen [a,1], Qianying Zheng [a,1], Jiansen Chen [b,*]

[a] College of Physics and Information Engineering, Fuzhou University, Fuzhou 350116, China
[b] Department of Nosocomial Infection Control, Fujian Medical University Union Hospital, Fuzhou 350001, Fujian, China

ARTICLE INFO

ABSTRACT

Objective: Medical image analysis is particularly important for doctors to differential diagnosis of diseases. Due to the outbreak of COVID-19, how to diagnose COVID-19 accurately has become a key issue. High-resolution lung CT images can provide more diagnostic information, so there is an urgent need to develop a super-resolution method to improve the resolution of medical images.

Methods: In this paper, a method based on double paths with residual information distillation for medical images super resolution (DRIDSR) is established. In the low-frequency path, shallow convolutional network is used to get low-frequency features, while in the high-frequency path, a residual information distillation module (RIDM) is designed to obtain clearer high-frequency features. RIDM cascades multiple residual blocks, and uses the output of each residual block as the input of IDB for further information distillation. Finally, it merges the information left by multiple IDBs as output.

Results: The proposed method is tested on the public dataset COVID-CT. The DRIDSR reconstruction quality of the algorithm is higher than that of the SRCNN, ESPCN, VDSR, IMDN and PAN method (+2.21 dB, +2.41 dB, +1.42 dB, +0.43 dB, +0.54 dB improvement, respectively) at × 3 upscale factor and (+2.35 dB, +2.17 dB, +1.59 dB, +0.48 dB, +0.56 dB increase, respectively) at ×4 upscale factor. While the number of parameters and analysis time of our model are reduced.

Conclusions: It is demonstrated that DRIDSR network can obtain better performance and better HR medical images than several state-of-the-art SR methods in terms of objective indicators and subjective evaluation.

1. Introduction

The ongoing outbreak of the novel coronavirus disease (COVID-19) that occurred in China is rapidly spreading globally. It has posed a threat to human lives, global economic security, and the healthcare system. And more scientific research is needed to put into the medical field. As the majority of patients infected with COVID-19 had lung computed tomography (CT) abnormality, many studies suggest lung CT should be a primary diagnostic tool for COVID-19. The higher the resolution of a lung CT image, the more detailed information it has. Therefore, doctor can diagnosis the disease accurately. The principle of image acquisition by medical equipment is complicated. In order to improve the resolution of medical images, many related complex technologies are involved, and more expensive equipment is required. The super-resolution (SR) algorithms can directly improve the resolution of existing medical images.

Without changing the current medical imaging equipment, the SR method can directly reconstruct high-resolution (HR) images from low-resolution (LR) images. The SR method can provide physician with more detailed and accurate diagnostic information. Since lung CT images are different from other natural images, the study of SR reconstruction algorithms for lung images is of great significance in the current medical diagnosis research fields.

Generally, SR algorithms can be divided into three main categories: i) the interpolation-based SR method, which obtain the pixel values of the HR image by non-uniform interpolation within a neighborhood of the pixel from the LR image. However, the methods can't obtain detailed HR images, and the reconstructed SR images are not satisfactory. ii) the reconstruction-based SR method [3,4,15], it can obtain clearer HR images than the interpolation-based SR method. However, when the magnification factor is increased, the high-frequency information

contained in the SR image generated by these methods will decrease. iii) The learning-based image SR methods, which utilize a large amount of HR-LR paired images to learn their mapping functions or explore the self-similarity between images. Although this method achieves better SR results than those of the previous two methods, it faces the difficulty of effective and compact modeling of the data. Method based on Markov network [6,12,19,20] is used to learn the relationship between high-frequency and low-frequency information of the image. And the method based on neighbor embedding [5,7,8,17] tries to obtain a mapping model for each patch to reconstruct the required SR image. Furthermore, sparse coding [13] and ridge regression methods have been adopted in the field of SR images. However, the disadvantage of sparsity-based techniques is that introducing sparsity constraints through nonlinear reconstruction is usually computationally expensive.

Recently, deep learning has played an advantage in computer vision tasks, such as image recognition, object detection and semantic segmentation. Dong et al. [10] first introduced a convolutional neural network (SRCNN) for a single image SR (SISR) and achieved a good result. Kim J et al. [14] introduced a residual structure [31] called VDSR into SISR. VDSR does not directly learn the mapping between low-resolution images and high-resolution images, but learns the residuals of the two types of images. The introduction of residual structure is a convergence speed of model training, and a deeper network structure is introduced into SISR, so that the model has a broader receiving field. Shi W et.al. [1] proposed an efficient sub-pixel convolutional neural network (ESPCN). The latter introduced an efficient sub-pixel convolution layer which learns an array of upscaling filters to upscale the final LR feature maps into the HR output. The method is widely used for subsequent instead of the transposed convolutional because of its effectiveness. However, ESPCN lacks contextual information, resulting in insufficient high-frequency information and insufficient edge details in the restored image. Zhang et.al. [25] drew on the channel attention mechanism of SENet [26] and proposed a Residual Channel Attention Networks (RCAN) to make the network focus on channels with richer features. In addition, RCAN uses long skip connection and short skip connection to make low-frequency information flow forward.

Although these works use a variety of methods to extract features through convolutional networks and obtain SR images, they do not consider distinguishing information between the low-frequency and the high-frequency of the image. For shallow layers, the parameters may be suitable for low-frequency information (with simple textures), but not suitable for high-frequency information (with complex textures). For deep layers, the parameters may be suitable for high-frequency information, but are over-fit for low-frequency information [32]. Separate reconstruction of low-frequency information and high-frequency information can obtain better SR results. The inconsistency between model complexity and frequency is a key issue that limits the performance of those deep CNN-based methods. Considering the above reasons and inspired by Information Multi-distillation Network (IMDN) [2], we propose a double paths network with residual information distillation super resolution (DRIDSR) for lung CT images. The main contributions of our work are summarized as follows:

(1). Introducing the residual information distillation module in the high frequency path way, which can predict refined high frequency information.
(2). DRIDSR is divided into two pathways, which are used for extract accurate low-frequency information and high-frequency information respectively.
(3). A new image gradient residual training method is proposed. By training the residuals of the image gradient features, the network can converge faster and obtain better high-frequency features.

2. Related work

In recent years, many image super-resolution methods based on deep

learning have been proposed [21–23]. Different from traditional interpolation methods, deep learning methods directly learn the mapping from LR images to HR images by constructing a network model. Due to the powerful learning ability of convolutional neural network, the deep learning based method has obtained perfect results.

The traditional SR method uses an iterative algorithm to generate HR images by exploring the prior information of the image. This method does not consider the relationship between the LR image and the HR image, and less information can be used, so the generated effect is less than ideal. As a pioneer of convolution neural network, SRCNN directly learns the nonlinear mapping relationship between LR images and HR images, and generates the desired HR images. The input of SRCNN is the bicubic interpolation of LR, which is called pre-upsampling. These pre-upsampling designs [10,14,12,16] alleviate the difficulty of learning, because the network only needs to learn the mapping of coarse input to HR. However, the pre-upsampling increases the model parameters and training time since the features are directly extracted in the HR space. More importantly, the output of the network will become fuzzy and lack useful high-frequency information.

To solve the problems caused by the above-mentioned pre-upsampling, many subsequent networks have adopted to the post-upsampling designs [18]. In contrast, the post-upsampling network uses LR as input, and performs an upsampling operation in the final reconstruction part of the network. The deconvolution layer [11] and sub-pixel convolution layer [9] are often used for image upsampling. The deconvolution layer has too many zero padding, making the result less than expected. However, sub-pixel convolution layer can obtain more information from the original image by rearranging the feature map, so the upsampling effect is better.

For medical image SR, Kang M S et al. [28] proposed a novel method that utilizes the SRCNN with image-based cell phenotype analysis to improve the quantification accuracy, and use automatic image processing to predict the response of glioblastoma cells to drugs. Qiu D et al. [28] proposed a method for efficient medical image super-resolution (EMISR). EMISR adopts a network structure combining SRCNN and ESPCN to achieve better SR results on knee magnetic resonance imaging (MRI) images. Wei Lu et al. [29] proposed a novel densely connected network for SR reconstruction of 3D medical images. The network utilizes a 3D dilated convolution module to increase the receptive field and obtain multi-scale information. Besides, a local residual dense attention module is introduced to learn rich features of the input 3D medical images. However, this method does not consider the potential correlation between 3D medical images.

3. Methods

3.1. Network architecture

As shown in Fig. 1, DRIDSR consists of two pathways: low-frequency information pathway and high-frequency information pathway. We denote $I_{LR} \in \mathbb{R}^{h \times w \times 1}$ as the input LR image and $I_{SR} \in \mathbb{R}^{sh \times sw \times 1}$ as the output SR image, where h and w represents the height and the width of the input image. s is the scale factor for super resolution. The low-frequency information pathway transfers low frequency information from LR images. The high-frequency information pathway extracts high frequency information from LR images by our RIDM. All Conv layer in our network use the same padding, so the size (h, w) of the input image has not changed. Finally, we add the outputs of the two pathways and upsample them through pixel shuffling. The network is optimized by minimizing the difference between HR image I_{HR} and SR image I_{SR}. Here, we use L1 loss function as our metrics, formulated as:

$$L = \frac{1}{n} \sum_{i=1}^{n} \left\| I_{SR}^i - I_{HR}^i \right\|_1, \tag{1}$$

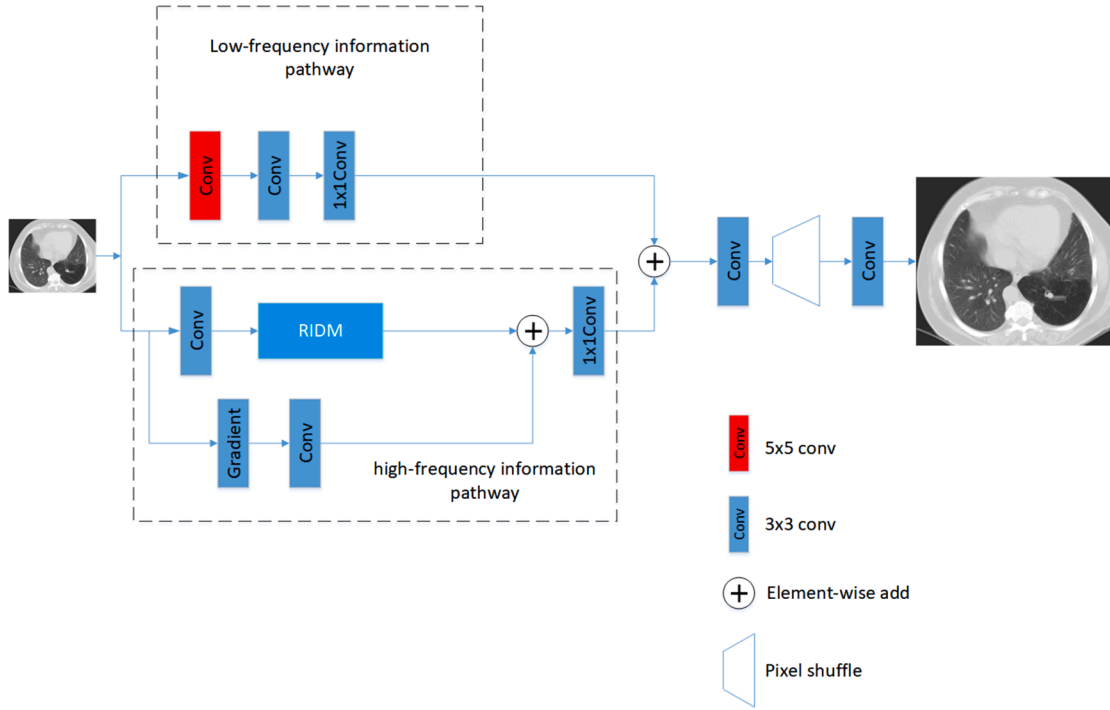where I_{HR} refers to the HR images, I_{SR} refers to the SR images, and n is

Fig. 1. The architecture of DRIDSR.

the batch size.

3.2. High-frequency information pathway

The LR image is fed into the high-frequency pathway. It is mainly composed of residual information distillation module (RIDM) and gradient identity path. Firstly, we use a Conv layer to extract the rough feature map of the input LR image. These feature maps will be fed into RIDM. For the gradient identity path, we use Sobel operator to obtain the edge map I_g of the image. As we all know, the Sobel operator is a commonly operator for extracting image gradients. We use the edge map I_g to guide the restoration of high-frequency information of the image. The goal of the gradient identity path is to learn the mapping from LR gradient map to HR map. The Sobel operator is a first-order differential operator. We obtain the Sobel gradient map of the image by calculating the difference between the center pixel and the surrounding eight pixels. The Sobel gradient map shows the gray-scale mutation area in the image, including high-frequency information such as edges and textures. The Sobel gradient map contains information that is severely missing during the SR reconstruction process. Therefore, fusing the information of the gradient identity path can effectively improve the quality of the SR image.

We utilize a Conv to extract the feature map F_g of the edge map. In order to train the high-frequency path efficiently, we add the output of RIDM and the edge feature map F_g together. This operation is similar to skipping connection, which can reduce the difficulty of training. Finally, we use a 1×1 Conv layer to reduce the number of features. This process can be written as:

$$F_M = f_{RIDM}(f_1(I_{LR})), \tag{2}$$

$$I_g = f_g(I_{LR}), \tag{3}$$

$$F_g = f_2(I_g), \tag{4}$$

$$F_H = f_3(F_M + F_g), \tag{5}$$

where f_1, f_2 and f_3 are the mapping of the convolution layer. f_{RIDM} is the

mapping of RIDM. f_g denotes the Sobel operator.

This gradient identity path enables the network to better learn the high-frequency information of the image, while the network only learns the high-frequency residual of the image, which can speed up the training speed of the network and make the network converge better. Based on RIDM, our proposed network can extract accurate high-frequency features and obtain significant SR results.

3.3. Residual information distillation module

The residual structure has shown great performance in image super-resolution. Referring to the information multi-distillation block (IMDB) [2], we propose a residual information distillation module (RIDM), as shown in Fig. 2. RIDM contains m information distillation blocks (IDBs) and n residual blocks. The features retained by each IDB and the features output by the last residual block are concatenated on the channel. With the fusion of features, the network will become larger and larger, it will be more difficult to train. Hence, we introduced a 1×1 Conv layer at the end of RIDM to control information. The process is defined as follows:

$$F_M = f_{1x1}([F_{r1}, F_{r2}, F_{r3}, F_{r4}, F_{r5}, F_6]) + F_0 \tag{6}$$

where $F_{r1}, F_{r2}, F_{r3}, F_{r4}, F_{r5}$ represent the refined feature maps produced by the IDB, and F_6 is the output of the final residual block. f_{1x1} represents the 1×1 convolution operation. F_M denotes the output of RIDM. Then, there is a local skip connection which can provide fast and stable training. By fusing the information refined by multiple IDBs, the network can capture the rich features for the final reconstruction.

RIDM uses an information distillation block (IDB) to extract more useful features from the output features of the residual block. For each IDB, we employ two convolutional operation on the previous features, which will produce two parts of features. One is preserved, and the other part is fed into the next IDB. The retaining part can be regarded as a refined features.

We use a residual block to extract deep features F_1, then take F_0 and F_1 as the inputs of the first IDB,

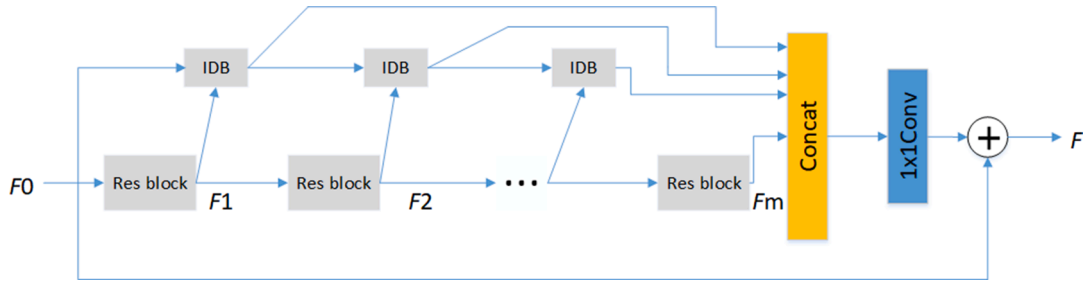$$F_{d1}, F_{r1} = f_{IDB}([F_0, F_1]), \tag{7}$$

Fig. 2. Residual information distillation module (RIDM).

where $[F_0, F_1]$ denotes the concatenation of F_0, and F_1. f_{IDB} denotes our proposed simple information distillation block. Starting from the second IDB, the input of each IDB is the output of the previous IDB and the output of the residual block. This process can be define as:

$$F_d^i, F_r^i = f_{IDB}\left(\left[F_d^{i-1}, F_i\right]\right), \quad i = 2, 3, \cdots, 6 \tag{8}$$

where $\left[F_d^{i-1}, F_i\right]$ represents the concatenation of the features produced by the (i-1)-th IDB and the (i)-th residual block. F_d^i and F_r^i denote the output of the (i)-th IDB.

3.4. Information distillation block

As shown in Fig. 3(a), the original information distillation block first uses a 3×3 convolution layer to extract input features for subsequent distillation steps. The main idea of this block is extracting useful features little by little like DenseNet. Then, the channel splitting operation is applied to the previous features and the input features are divided into two parts. One part is retained and the other part is fed into the next distillation step. Given the input features X_1, X_2, this procedure can be described as:

$$F_d, F_r = \text{Split}(f_{3x3}([X_1, X_2])) \tag{9}$$

where F_d and F_r denote the coarse features and refined features. $f_{3\times3}$ denotes the 3×3 Conv layer. Split denotes the channel splitting operation.

Although the origin IDB has achieved prominent improvements, it is not efficient enough and introduces some inflexibility due to the channel splitting operation. The coarse features are generated by 3×3 convolution filter that has many redundant parameters. Based on the above considerations, we redefined the simple information distillation block (IDB) as Fig. 3(b) shown:
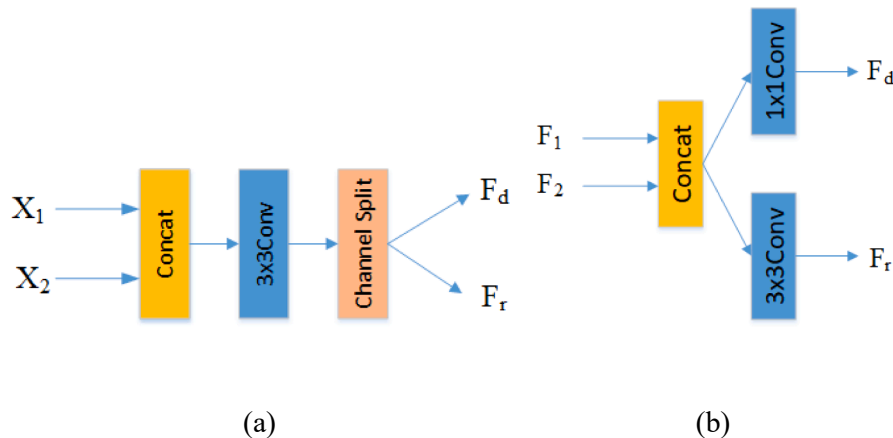
$$F_d = f_{1x1}([X_1, X_2]), \tag{10}$$

$$F_r = f_{3x3}([X_1, X_2]), \tag{11}$$

where X_1 and X_2 denote the inputs of IDB, F_d and F_r denote the coarse features and refined features. $f_{1\times1}$ and $f_{3\times3}$ denote the 1×1 Conv and 3×3 Conv, respectively. This simple structure can filter information effectively. In addition, it is more flexible than the original IDB and improves the network performance.

3.5. Low-frequency information pathway

In VDSR, the network uses skip connections to add the input and output images together. The network predicts the residual image between the LR and HR images. Similarly, we designed a low-frequency information pathway. However, in order to increase the high-frequency information pathway, we use two convolutional layers to increase the number of feature channels. Then we use a 1×1 Conv to adjust the number of feature channels to match the output of the high-frequency information pathway. It is formulated as follow:

$$F_L = f_{1\times1}(f_{3\times3}(f_{5\times5}(I_{LR}))) \tag{12}$$

where F_L represents the output features of the low-frequency path. $f_{1\times1}$, $f_{3\times3}$ and $f_{5\times5}$ denote convolutional layers with different kernel size.

Then, we add the output of the low-frequency path and the high-frequency path to merge the information of the two paths and to reconstruct a clear and undistorted medical SR image. Finally, the features are transferred to the reconstruct module to obtain the final SR image. This process can be written as:

$$F = F_L + F_H, \tag{13}$$

where F represents the final merged features of the high-frequency path and the low-frequency path.

(a)	(b)

Fig. 3. (a) the original information distillation block, and (b) our redesigned information distillation block.

3.6. Reconstruct module

In ESPCN, the network utilizes sub-pixel convolution layers to directly generate HR images from LR feature maps. Due to its advantage, pixel shuffling has always been a common layer of upsampling features. We use pixel shuffling to upsample the features and get the final output. Firstly, we use a Conv layer to the output F to further merge the high and low frequencies and reduce the number of feature maps. It takes C × H × W low resolution images as an input and outputs s^2 × C × H × W features, and s denotes upscale factor. Secondly, the features are upsampled by pixel shuffling [1]. Finally, a 3 × 3 convolutional layer changes the number of output channels to 3. This process can be written as:

$$I_{SR} = f_{3\times3}(f_{ps}(f_{3\times3}(F))) \tag{14}$$

where $f_{3\times3}$ is 3 ×3 Conv layer and f_{ps} denotes pixel shuffling operation.

3.6.1. Sub-pixel convolution layer

Pixel shuffle is an up-sampling method that can effectively enlarge the feature map. It can replace interpolation or deconvolution method to achieve upscaling. The main purpose of pixel shuffle (pixel reorganization) is to obtain high-resolution feature maps from low-resolution feature maps through convolution and multi-channel recombination. The sub-pixel upsampler is based on the depth information of feature map, which can effectively improve the up-sampling quality.

As shown in Fig. 4, pixel shuffle is an periodic shuffling operator that rearranges the elements of a $H \times W \times C \cdot r^2$ tensor to a tensor of shape $rH \times rW \times C$. Mathematically, this operation can be described in the following way

$$\mathscr{PS}(\mathrm{T})_{x,y,c} = \mathrm{T}_{\lfloor x/r \rfloor, \lfloor y/r \rfloor, C \cdot r \cdot \bmod (y,r) + C \cdot \bmod (x,r) + c}$$

where PS represent the operator of pixel shuffle. And mod(x,r), mod(y,r) are the different sub-pixel locations, where x,y are the output pixel coordinates in HR space.

4. Experiments

The experimental environment consists of a workstation (Intel® Core™ i7-7700 CPU @ 3.60 GHz) with GPU NVIDIA GeForce GTX 1080 Ti, and software including CUDA Toolkit v10.1 and Pycharm. The Pytorch framework is built for training and testing.

4.1. Datasets

Most current SR models are all trained on the DIV2K [27] dataset. However, DIV2K is mainly contains natural images, which is quite different from medical CT images. The SR results of a model trained on DIV2K for medical CT images may not be suitable.

Considering that our model is applied for the SR reconstruction of medical images, we use the public COVID-CT dataset [30], which contains 349 COVID-19 CT images from 216 patients and 397 non-COVID-19 CT images. In the division of the dataset, we used 600 images for training, 100 images were verified in training, and finally 46 images were tested.

4.2. Implementation details

In view of the common magnification in the clinical medical field, we only trained the SR network with an up-scale factor m (m = 3, 4). For preparing the input LR images, we downsample the original HR image with an upscaling factor m by using a bicubic downsampling operation to generate the corresponding LR image. We crop 192 × 192 patches from the HR images. Then according to the upscaling factor m, the corresponding LR patches size is $\frac{192}{m} \times \frac{192}{m}$. We obtained 1498 LR-HR patch pairs. The batch size is set to 64. For data augmentation, we perform randomly horizontal flip and 90°, 180°, 270° rotation. We set the kernel size to 3 × 3, and the stride and padding to 1 to maintain the feature maps size of the convolutional layers unchanged. For 5x5 Conv, we set the padding to 2. All activations in our network are ReLU. We use Adam optimizer [19], where $\beta_1 = 0.9$, $\beta_2 = 0.999$, and $\epsilon = 10^{-8}$. The learning rate is initially set to 1×10^{-3} and halved at every 30 epochs. The final learning rate is 1×10^{-5}, and we stop training after 200 epochs.

For the gradient identity path in the high-frequency path, we use the Sobel operator, a simple and fast method to obtain the edge map of the image. Because the texture information of CT images is relatively simple, there is no need to use the complex edge detector such as Canny [39]. Moreover, Canny relies on the threshold parameter, which affects the performance of the high-frequency path. As shown in Fig. 5, it is demonstrated that the edge map generated by Sobel retains the primary edge of the image. While the edge map obtained by Canny has some redundant edges. The threshold of Canny is set to 60 (minval) and 120 (maxval).

Similar to the report by Ren et al. [38], we use the peak signal-to-noise ratio (PSNR), structural similarity index (SSIM), multi-scale SSIM (MS-SSIM) and perceptual index (PI) as performance metrics to evaluate our model. PSNR and SSIM metrics values are evaluated on the Y channel for testing. PSNR is an image quality evaluation index based on the error sensitivity between corresponding pixels. PSNR is the most commonly and most widely used objective evaluation index for images, as shown in formulas 15 and 16 [40].The SSIM index is used to measure the similarity of two images. SSIM measures similarity of images from three aspects: brightness, contrast, and structure, as shown in formula 17. In this method, X is defined as SR, Y represents ground truth image HR, H and W are the height and width of the image, and X(i,j) and Y(i,j) represent (i, j) pixels in SR and HR. MS-SSIM is a multi-scale similarity index, which is closer to the results of subjective visual evaluation than SSIM. PSNR and SSIM metrics are widely used in the natural image and medical image SR evaluation [33,34,38]. The effectiveness of PSNR and SSIM is unclear, and the required ground-truth images are not always available in practice. To address the issue and evaluate the quality of image itself, we add the non-reference image quality assessment metric PI. The PI [37] is judged by the score of Ma [35] and the non-reference metric of NIQE [36], as shown in formula 18. PI has been proven to be
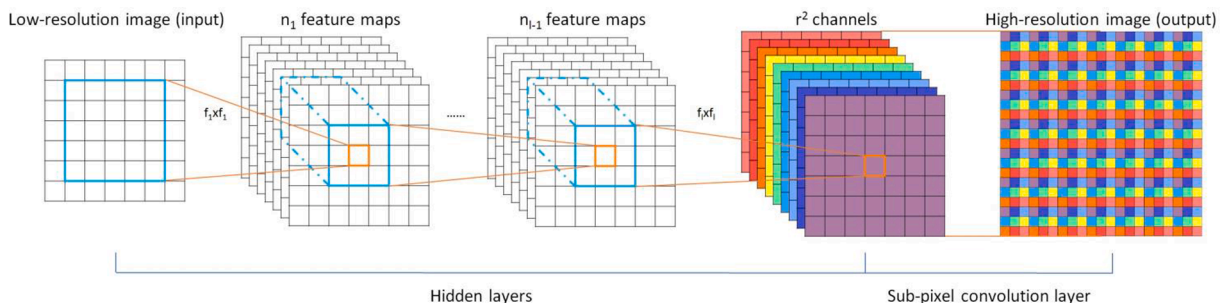

Fig. 4. The efficient sub-pixel convolutional neural network (ESPCN) and the sub-pixel convolution layer.
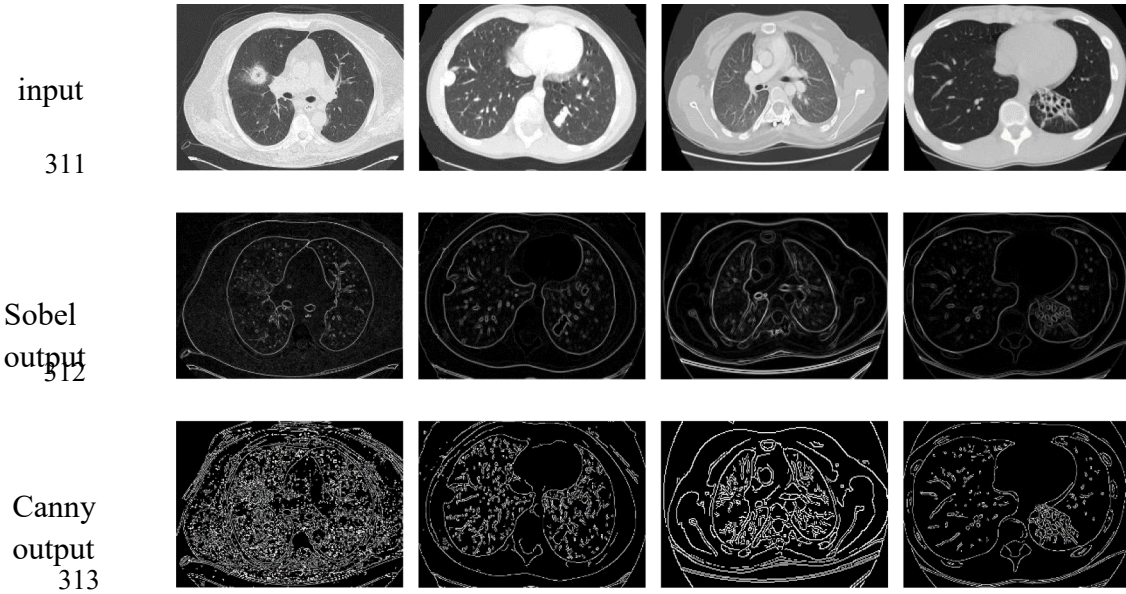
Fig. 5. The edge maps generated by Sobel and Canny, respectively.

highly correlated with human ratings. The lower PI value means better results.

$$MSE = \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} (X(i,j) - Y(i,j))^2 \qquad (15)$$

$$\sum_{j=1}^{W} (X(i,j) - Y(i,j))^2 \qquad (16)$$

$$SSIM(X,Y) = l(X,Y) \cdot c(X,Y) \cdot s(X,Y) \qquad (17)$$

$$PI = \frac{1}{2}((10 - \text{Ma}) + NIQE) \qquad (18)$$

4.3. Image super resolution results

In this study, we used a double-path network to achieve super-resolution of high-frequency and low-frequency respectively. As shown in Fig. 6, the output of the low-frequency path retains the rough outline (low-frequency features) of the input image, and the output of high-frequency remain the texture and detail information of the input image. It is revealed that the proposed method can significantly separate the high and low frequency information of the input image. Therefore, our network can obtain better SR result through the frequency separation mechanism.

Compared with all previous methods, our DRIDSR performs better than other methods on all datasets with all scaling factors. We show the visual results compared with other benchmarking methods with a ×3 upscale factor. According to the display results, it is shown that the image reconstructed by DRIDSR is clearer at the edge of the lungs and small nodules in the lungs (Figs. 8, 9). In Fig. 7, it is shown that the super-resolution results of images with Ground-Glass Opacity (GGO) at ×4 upscale factor. GGO is a sign of the lungs on high-resolution CT. It is manifested as a slight increase in lung CT density, and the outlines of bronchi and blood vessels can be seen, similar to ground glass. Compared with other methods, the GGO in our SR results has clearer outline and details. Due to the gradient residual training method, this method can obtain clearer GGO, and can be clearly observed the pathological features in lung CT images. (Figs. 10, 11)

PSNR and SSIM indexes are important performance indicators for SR reconstruction. As shown in Table 1, our DRIDSR has higher PSNR and SSIM, which shows that the quality of SR results is better than other SOTA methods and more similar to HR images. In addition, the proposed method achieves the highest MS-SSIM and the lowest PI. The results show that our model has a greater progress in visual perception. After the SR reconstruction, the details of the lungs are clearly visible, which can effectively help doctor make an accurate diagnosis.

Then, we compare the parameters of the VDSR, IMDN, PAN and our DRIDSR model, since these methods have similar scale of parameters, as shown in Table 2. Based on the results, the following conclusions can be
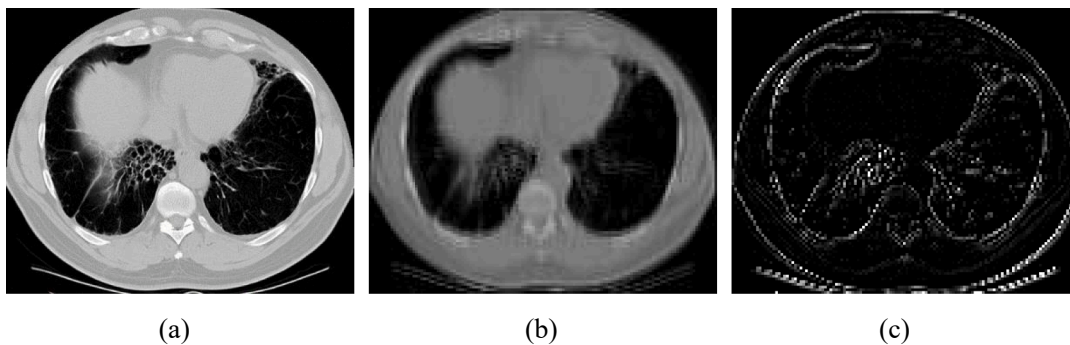

(a)   (b)   (c)

Fig. 6. The output feature map of low-frequency path and high-frequency path. (a) the input image (b) the output feature map of low-frequency path (c) the output feature map of high-frequency path.
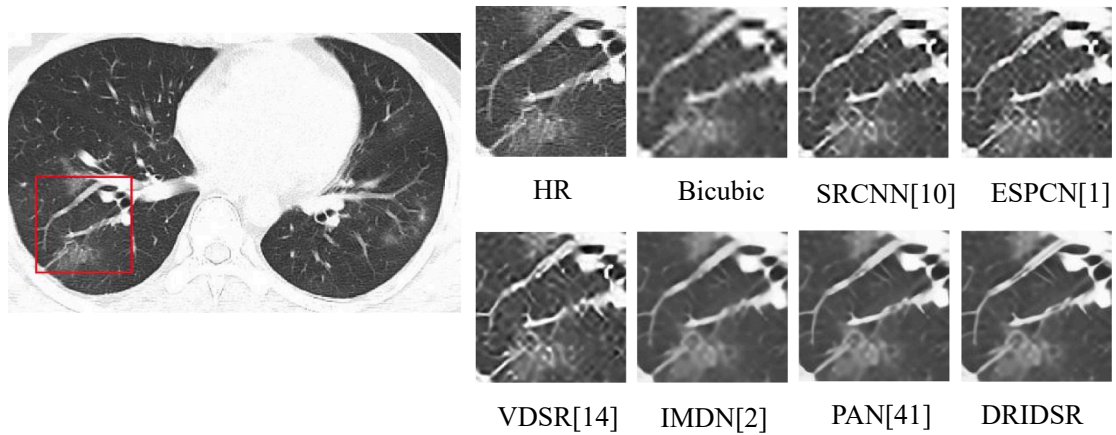
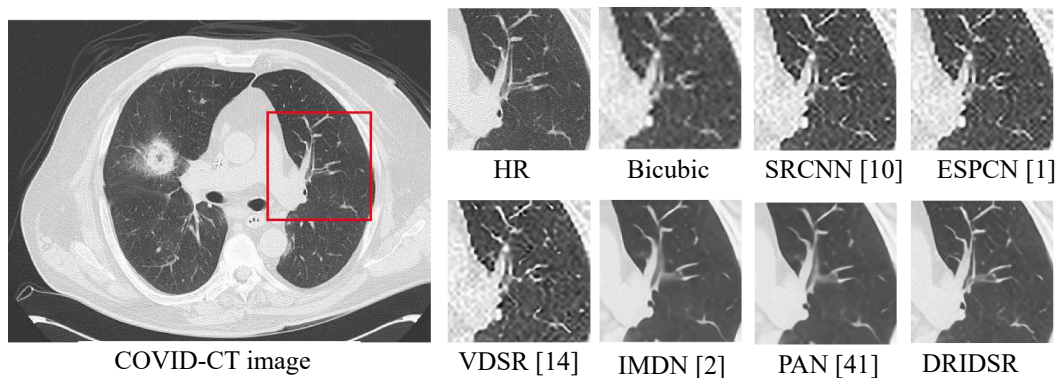**Fig. 7.** Visual comparison for ×4 SR on GGO image patch.

Fig. 8. Visual comparison for ×3 SR on a COVID-CT test image.

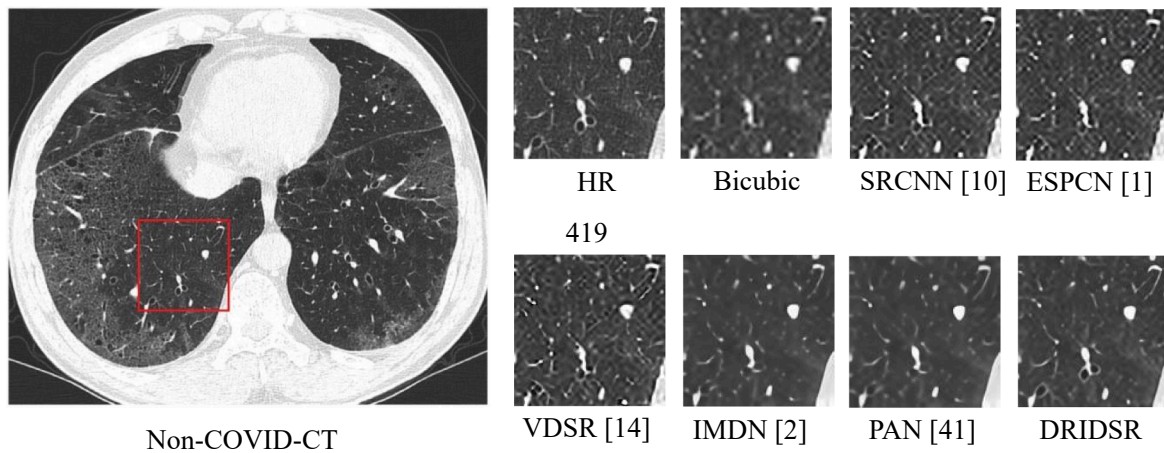**Fig. 9.** Visual comparison for ×3 SR on Non-COVID-CT test image.

drawn. Although our DRIDSR model has fewer parameters than VDSR and IMDN (Second only to PAN), the SR results are much better.

We also test the running time during the feed-forward process. It can be seen from Table 2 that our DRIDSR has a faster inference speed while maintaining excellent SR performance.

5. Discussion

The deep learning algorithm uses offline training, so the trained model can be directly used for image processing. Therefore, "deep learning" methods can achieve faster reconstruction speed while guaranteeing reconstruction quality.

Compared with natural images, the super-resolution of CT images is relatively simple. CT images are usually gray-scale images with obvious edges fewer detailed features that need to be restored. We believe that the super-resolution of CT images does not require too complex models, and often simple models can achieve good results. Compared with other low parameters models such as VDSR and IMDN, our DRIDSR has only 0.52 M and obtain the best SR performance. Although PAN has lowest parameters, the edge details of the image are not sharp enough.
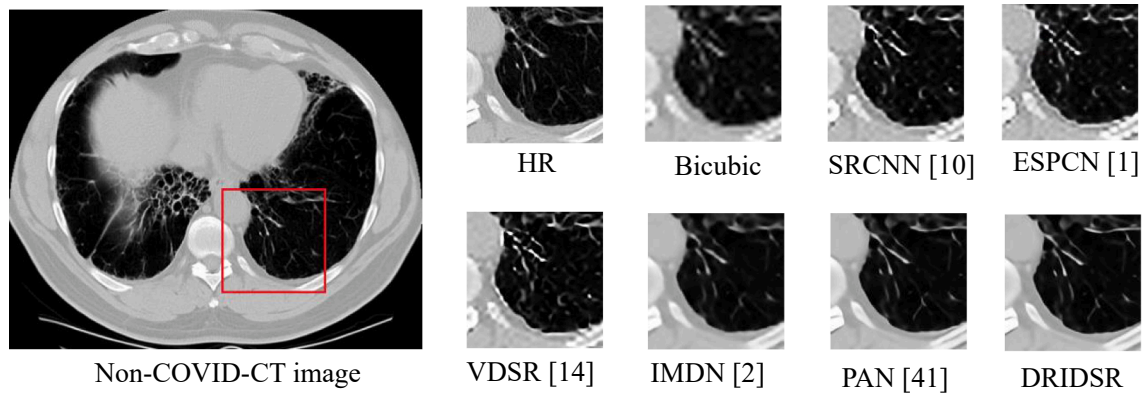
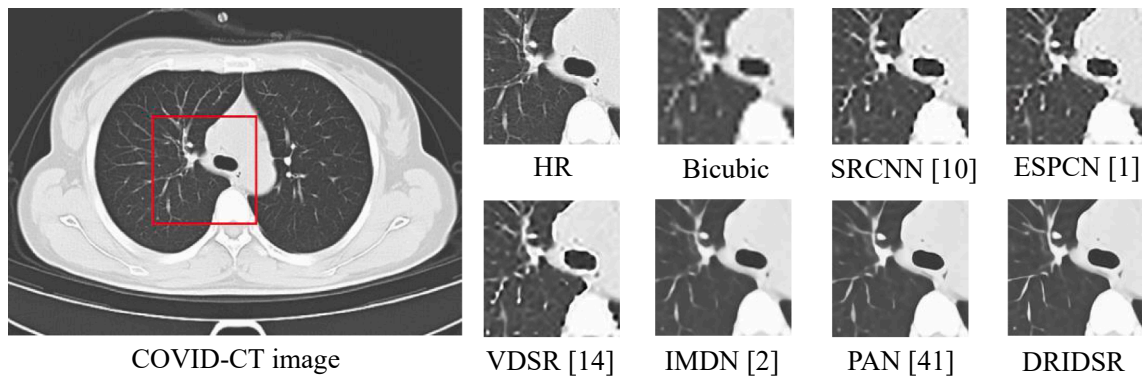**Fig. 10.** Visual comparison for ×4 SR on a Non-COVID-CT test image.

Fig. 11. Visual comparison for ×4 SR on a COVID-CT test image.

Table 1
Comparison of PSNR, SSIM between DRIDSR and several advanced methods in detection of COVID-CT images.

Method	Scale	COVID-CT			
		PSNR	SSIM	MS-SSIM	PI
Bicubic	×3	29.11	0.8274	0.9619	9.1833
SRCNN [10]	×3	32.26	0.8760	0.9763	8.2775
ESPCN [1]	×3	32.06	0.8723	0.9755	8.0422
VDSR [14]	×3	33.05	0.8861	0.9788	7.9292
IMDN [2]	×3	34.04	0.8839	0.9799	7.2477
PAN [41]	×3	33.93	0.8848	0.9798	7.3347
DRIDSR(ours)	×3	**34.47**	**0.8874**	**0.9805**	**7.1783**
Bicubic	×4	26.85	0.7550	0.9369	9.4331
SRCNN [10]	×4	29.54	0.8152	0.9591	8.7943
ESPCN [1]	×4	29.72	0.8176	0.9601	8.5870
VDSR [14]	×4	30.30	0.8343	0.9644	8.6821
IMDN [2]	×4	31.41	0.8348	0.9696	7.8519
PAN [41]	×4	31.33	0.8360	0.9699	7.9969
DRIDSR(ours)	×4	**31.89**	**0.8406**	**0.9714**	**7.8248**

Table 2
The comparison of model parameters and performance with VDSR and DRIDSR for scale factor ×4 in COVID-19 test dataset.

Methods	Parameters (M)	Layers	PSNR	SSIM	Time (s)
VDSR [14]	0.67	**20**	30.30	0.8343	0.0825
IMDN [2]	0.71	34	31.41	0.8348	0.0638
PAN [41]	**0.27**	60	31.33	0.8360	**0.0427**
DRIDSR	0.52	32	**31.89**	**0.8370**	0.0548

It is worth nothing that the experiments are conducted on a COVID-19 lung CT image dataset. All the experimental results show that the proposed DRIDSR algorithm outperforms all the compared algorithms, including the state-of-the-art IMDN and PAN algorithms. We compared the three super-resolution reconstruction methods by learning, SRCNN [10], ESPCN [1], VDSR [14], IMDN [2], PAN [41] and a traditional interpolation-based function, Bicubic. The objective metrics such as PSNR, SSIM, MS-SSIM and PI are evaluated, and the images were also evaluated subjectively, which is commonly used in existing studies [24,28,29,34,38]. For Fig. 8 and Fig. 10, the SR result generated by our DRIDSR has a clearer edge of lung. Besides, the soft tissues in other parts appear clearer. The reason why our DRIDSR is good for edge information recovery is that we use the edge map of the image to constrain the high-frequency path. As shown in Fig. 9 and Fig. 11, compared with other methods, our method has better SR performance for small nodules in the lung. This is due to the following two reasons. First, the DRIDSR has two paths (low-frequency path and high-frequency path), and each path focus on the SR of specific information. Second, in the high-frequency path, we use the interaction between the residual block and the IDB, so that the high-frequency path can refine the high-frequency information better.

Compared with other methods, our DRIDSR achieves the best PSNR, SSIM, MS-SSIM and PI, which means the SR image generated by DRIDSR has the best quality and visual perception. The results show that the proposed method can not only improve the resolution of lung CT images, but also retain more medical and pathological characteristics. Lung CT images with rich pathological characteristics will be beneficial to doctors' diagnosis.

6. Conclusions

For the super-resolution of lung CT images, the main problem is that the model lacks the ability to recover high-frequency information. In this paper, we proposed a double path network with residual information distillation for medical image super resolution. The proposed network is divided into double paths, which are used for accurately extract the low-frequency information and high-frequency information of the image, respectively. And a novel image gradient residual training method is proposed. By training the residuals of the image gradient features, the network can converge faster and obtain better high-frequency features. The residual information distillation module is introduced into high-frequency path, which can predicts the refined high-frequency information. The proposed method achieves excellent performance on COVID-CT dataset. In the future, we will carry out research work from three aspects. i) introduce blind SR, which considers the real degenerative nucleus of the lung CT image for super-resolution. ii) take different super-resolution methods for the lesion area and non-lesion area of the image. And iii) find a better loss function to achieve better SR for high-frequency and low-frequency information, respectively.

Ethical approval

No ethics approval is required.

CRediT authorship contribution statement

Yihan Chen: Conceptualization, Methodology, Investigation. **Qianying Zheng:** Software, Validation, Writing – original draft. **Jiansen Chen:** Writing – review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

References

[1] W. Shi, J. Caballero, Ferenc Huszár, et al., Real-Time Single Image and Video Super-Resolution Using an Efficient Sub-Pixel Convolutional Neural Network. CVPR 2016, IEEE, 2016.

[2] Z. Hui, X. Gao, Y. Yang, et al. Lightweight Image Super-Resolution with Information Multi-distillation Network. 2019.

[3] M. Protter, M. Elad, H. Takeda, P. Milanfar, Generalizing the nonlocal-means to super-resolution reconstruction, IEEE Trans. Image Process 18 (1) (2009) 36–51.

[4] F. Rousseau, A non-local approach for image super-resolution using intermodality priors, Med. Image Anal. 14 (4) (2010) 594–605.

[5] M. Bevilacqua, A. Roumy, C. Guillemot, M.-L.-A. Morel, Neighbor embedding based single-image super-resolution using semi-nonnegative matrix factorization, in: 2012 IEEE international conference on acoustics, speech and signal processing (ICASSP), IEEE, 2012, pp. 1289–1292.

[6] R. Timofte, V. De Smet, L. Van Gool, Anchored neighborhood regression for fast example-based super-resolution, in: Proceedings of the IEEE international conference on computer vision, 2013, pp. 1920–1927.

[7] H. Chang, D.-Y. Yeung, Y. Xiong, Super-resolution through neighbor embedding. Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, IEEE, 2004. CVPR 2004, pp I–I.

[8] X. Chen, C. Qi, Low-rank neighbor embedding for single image super-resolution, IEEE Signal Process. Lett. 21 (1) (2013) 79–82.

[9] W. Shi, J. Caballero, F. Huszár, J. Totz, A.P. Aitken, R. Bishop, D. Rueckert, Z. Wang, Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network, in: Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 1874–1883.

[10] C. Dong, C.C. Loy, K. He, X. Tang, Learning a deep convolutional network for image super-resolution, in: European conference on computer vision, Springer, 2014, pp. 184–199.

[11] C. Dong, C.C. Loy, X. Tang, Accelerating the super-resolution convolutional neural network, in: European conference on computer vision, Springer, 2016, pp. 391–407.

[12] J. Kim, J. Kwon Lee, K. Mu Lee, Deeply-recursive convolutional network for image super-resolution, in: Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 1637–1645.

[13] J. Yang, J. Wright, T.S. Huang, Y. Ma, Image super-resolution via sparse representation, IEEE Trans. Image Process. 19 (11) (2010) 2861–2873.

[14] J. Kim, J. Kwon Lee, K. Mu Lee, Accurate image super-resolution using very deep convolutional networks, in: Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 1646–1654.

[15] J.-B. Huang, A. Singh, N. Ahuja, Single image super-resolution from transformed self-exemplars, in: Proceedings of the IEEE conference on computer vision and pattern recognition, 2015, pp. 5197–5206.

[16] Y. Tai, J. Yang, X. Liu, Image super-resolution via deep recursive residual network, in: Proceedings of the IEEE conference on computer vision and pattern recognition, 2017, pp. 3147–3155.

[17] K. Zhang, X. Gao, X. Li, D. Tao, Partially supervised neighbor embedding for example-based image super-resolution, IEEE J. Sel. Top. Signal. Process. 5 (2) (2010) 230–239.

[18] R. Timofte, E. Agustsson, L.V. Gool, M. Yang, L. Zhang, Ntire 2017 challenge on single image super-resolution: methods and results, in: IEEE Conf. Comput. Vis. Pattern Recognit. Workshops, 2017, pp. 1110–1121.

[19] D.P. Kingma, J. Ba. Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980, 2014.

[20] D. Martin, C. Fowlkes, D. Tal, J. Malik, A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics, in: Proc. 8th Int'l Conf. Computer Vision, 2001, pp. 416–423.

[21] J. Kim, J.K. Lee, L.K.M. Deeply-Recursive Deeply-Recursive Convolutional Network for Image Super-Resolution[J]. 2015.

[22] Y. Tai, J. Yang, X. Liu, Image super-resolution via deep recursive residual network. IEEE Conference on Computer Vision & Pattern Recognition, IEEE, 2017.

[23] W.S. Lai, J.B. Huang, N. Ahuja, et al., Deep Laplacian Pyramid Networks for Fast and Accurate Super-Resolution, in: IEEE Conference on Computer Vision & Pattern Recognition, IEEE Computer Society, 2017, pp. 5835–5843.

[24] J. Li, F. Fang, K. Mei, et al., Multi-scale Residual Network for Image Super-Resolution. 15th European Conference, Munich, Germany, September 8-14, 2018, Proceedings Part VIII[J], 2018.

[25] Y. Zhang, K. Li, K. Li, et al. Image Super-Resolution Using Very Deep Residual Channel Attention Networks, 2018.

[26] J. Hu, L. Shen, S. Albanie, et al., Squeeze-and-excitation networks, IEEE Trans. Pattern Anal. Mach. Intelligence (2017) 99.

[27] R. Timofte, E. Agustsson, L. Van Gool, M.H. Yang, L. Zhang, B. Lim, S. Son, H. Kim, S. Nah, K.M. Lee, et al., Ntire 2017 challenge on single image super-resolution: Methods and results. CVPRW, 2017.

[28] M.-S. Kang, E. Cha, E. Kang, J.C. Ye, N.-G. Her, J.-W. Oh, D.-H. Nam, M.-H. Kim, S. Yang, Accuracy improvement of quantification information using super-resolution with convolutional neural network for microscopy images, Biomed. Signal Process. Control 58 (2020) 101846, https://doi.org/10.1016/j.bspc.2020.101846.

[29] W. Lu, Z. Song, J. Chu, A novel 3D medical image super-resolution method based on densely connected network, Biomed. Signal Process. Control 62 (2020) 102120, https://doi.org/10.1016/j.bspc.2020.102120.

[30] X. Yang, X. He, J. Zhao, et al. COVID-CT-Dataset: A CT Scan Dataset about COVID-19[J]. 2020. [Online]. Available: https://www.graviti.cn/open-datasets /COVID_CT.

[31] K. He, X. Zhang, S. Ren, et al., Deep Residual Learning for Image Recognition. IEEE Conference on Computer Vision & Pattern Recognition, IEEE Computer Society, 2016.

[32] Y. Qiu, R. Wang, D. Tao, et al., Embedded block residual network: a recursive restoration model for single-image super-resolution, in: Proceedings of the IEEE/ CVF International Conference on Computer Vision, 2019, pp. 4180–4189.

[33] Q. Lyu, H. Shan, G.e. Wang, MRI super-resolution with ensemble learning and complementary priors, IEEE Trans. Comput. Imaging 6 (2020) 615–624.

[34] Y. Li, B. Sixou, F. Peyrin, A review of the deep learning methods for medical images super resolution problems, IRBM 42 (2) (2021) 120–133.

[35] C. Ma, C.-Y. Yang, X. Yang, M.-H. Yang, 'Learning a no-reference quality metric for single-image super-resolution', Comput. Vis. Image Understand. 158 (2017) 1–16.

[36] A. Mittal, R. Soundararajan, A.C. Bovik, 'Making a 'completely blind' image quality analyzer', IEEE Signal Process. Lett. 20 (3) (2013) 209–212.

[37] Y. Blau, R. Mechrez, R. Timofte, The PIRM challenge on perceptual image super-resolution, Proc. Eur. Conf. Comput. Vis. 2018 (2018) 1–22.

[38] S. Ren, J. Li, K. Guo, F. Li, Medical video super-resolution based on asymmetric back-projection network with multilevel error feedback, IEEE Access 9 (2021) 17909–17920.

[39] J. Canny, A computational approach to edge detection, IEEE Trans. Pattern Anal. Mach. (6) (1986) 679–698.

[40] A. Hore, D. Ziou, Image quality metrics: PSNR vs. SSIM, in: Proc. 20th Int. Conf. Pattern Recognit., 2010, pp. 2366–2369.

[41] H. Zhao, X. Kong, J. He, et al., Efficient image super-resolution using pixel attention, in: European Conference on Computer Vision, Springer, Cham, 2020, pp. 56–72.